

National Library
of Medicine

PubMed

PubMed

Nucleotide

Protein

Genome

Structure

PopSet

Taxonomy

OMIM

Search

PubMed

 for Go Clear Limits Preview/Index History Clipboard Display Abstract Save Text Order Add to ClipboardJ 1: *Mol Biol Evol* 1998 Jan;15(1):17-27 Details Citation Email

The evolutionary relationships between the two bacteria *Escherichia coli* and *Haemophilus influenzae* and their putative last common ancestor.

de Rosa R, Labedan B

Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay, France.

We have tried to approach the nature of the last common ancestor to *Haemophilus influenzae* and *Escherichia coli* and to determine how each bacterium could have diverged from this putative organism. The approach used was exhaustive analysis of the homologous proteins coded by genes present in these bacteria, using as criteria for sequence relatedness an alignment of at least 80 amino acid residues and a PAM distance (number of accepted point mutations per 100 residues separating two sequences) below 250. Evolutionarily significant similarities were found between 1,345 *H. influenzae* proteins (85% of the total genome) and 3,058 *E. coli* proteins (75% of the total genome), many of them belonging to families of various sizes (from 666 doublets to 35 large groups of more than 10 members). Nearly all the genes found by this approach to be duplicated in both bacteria were already duplicated in their last common ancestor. This was deduced from (1) the comparison of the respective distributions of evolutionary distances between orthologs (genes separated only by speciation events) and paralogs (genes duplicated in the same genome) and (2) the analysis of the phylogenetic trees reconstructed for each family of paralogs containing at least two members belonging to each bacterium. The distributions of the different categories of homologs show a significant loss of paralogous genes in *H. influenzae* (reduction proportional to the genome size), of many sequences which are still present in one copy in *E. coli*, and of some entire gene families. Phylogenetic trees also confirmed this recent loss of paralogous genes in *H. influenzae*. Thus, the genome size of the last common ancestor of these two bacteria would have been close to that of present-day *E. coli*, and the evolution of *H. influenzae* toward a parasitic life led to an important decrease in its genome size by some mechanism of streamlining. During this recent evolution, the memory of the gene order